

Genetics and population analysis

AncestryGrapher toolkit: Python command-line pipelines to visualize global- and local- ancestry inferences from the RFMIX version 2 software

Alessandro Lisi^{1,*} and Michael C. Campbell  ^{1,*}

¹Department of Biological Sciences (Human and Evolutionary Biology Section), University of Southern California, Los Angeles, CA 90089, United States

*Corresponding authors. Department of Biological Sciences (Human and Evolutionary Biology Section), University of Southern California, Los Angeles, CA 90089, United States. E-mails: alisi@usc.edu (A.L.) and mc44680@usc.edu (M.C.C.)

Associate Editor: Christina Kendziora

Abstract

Summary: Admixture is a fundamental process that has shaped levels and patterns of genetic variation in human populations. RFMIX version 2 (RFMIX2) utilizes a robust modeling approach to identify the genetic ancestries in admixed populations. However, this software does not have a built-in method to visually summarize the results of analyses. Here, we introduce the AncestryGrapher toolkit, which converts the numerical output of RFMIX2 into graphical representations of global and local ancestry (i.e. the per-individual ancestry components and the genetic ancestry along chromosomes, respectively).

Results: To demonstrate the utility of our methods, we applied the AncestryGrapher toolkit to visualize the global and local ancestry of individuals in the North African Mozabite Berber population from the Human Genome Diversity Panel. Our results showed that the Mozabite Berbers derived their ancestry from the Middle East, Europe, and sub-Saharan Africa (global ancestry). We also found that the population origin of ancestry varied considerably along chromosomes (local ancestry). For example, we observed variance in local ancestry in the genomic region on Chromosome 2 containing the regulatory sequence in the *MCM6* gene associated with lactase persistence, a human trait tied to the cultural development of adult milk consumption. Overall, the AncestryGrapher toolkit facilitates the exploration, interpretation, and reporting of ancestry patterns in human populations.

Availability and implementation: The AncestryGrapher toolkit is free and open source on <https://github.com/alisi1989/RFmix2-Pipeline-to-plot>.

1 Introduction

Admixture, defined as gene flow between previously diverged source populations leading to new populations with ancestry from multiple sources, has been an important contributor to levels and patterns of genetic variation in humans (Rudan 2006, 1000 Genomes Project Consortium 2012, Hellenthal *et al.* 2014, Gurdasani *et al.* 2015, Uren *et al.* 2016, Korunes and Goldberg 2021). For example, a number of studies have shown that regions of the modern human genome have originated from archaic hominins, such as Neanderthals and Denisovans (Qin and Stoneking 2015, Sánchez-Quinto and Lalueza-Fox 2015, Sankararaman *et al.* 2016, Wall and Yoshihara Caldeira Brandt 2016, Browning *et al.* 2018, Vyas and Mulligan 2019, Koller *et al.* 2022, Vespasiani *et al.* 2022, Witt *et al.* 2023). Other analyses have also revealed that admixture among modern human populations has occurred over millennia, leading to complex patterns of genomic diversity (Campbell and Tishkoff 2008, Campbell *et al.* 2014, Bekada *et al.* 2015, Arauna *et al.* 2017, Korunes and Goldberg 2021, Gopalan *et al.* 2022).

Admixture in populations can be inferred by examining genetic ancestry at two levels: global and local (Tan and

Atkinson 2023). Global ancestry refers to the proportion of an individual's genome that originated from distinct source populations, while local ancestry refers to an individual's genetic ancestry from source populations at particular chromosomal locations (Liu *et al.* 2013, Goli *et al.* 2024). To characterize patterns of admixture, several computational methods have been developed to infer global ancestry (Pritchard *et al.* 2000, Hoggart *et al.* 2004, Patterson *et al.* 2004, Alexander *et al.* 2009, Lawson *et al.* 2012, Liu *et al.* 2013, Hellenthal *et al.* 2014, Raj *et al.* 2014, Wangkumhang *et al.* 2022) and local ancestry along chromosomes in individuals (Falush *et al.* 2003, Tang *et al.* 2005, 2006, Sankararaman *et al.* 2008a, 2008b, Sundquist *et al.* 2008a, 2008b; Pasaniuc *et al.* 2009, Price *et al.* 2009, Bryc *et al.* 2010, Baran *et al.* 2012, Brisbin *et al.* 2012, Churchhouse and Marchini 2013, Maples *et al.* 2013, Durand *et al.* 2014, Guan 2014, Dias-Alves *et al.* 2018, Salter-Townshend and Myers 2019, Montserrat *et al.* 2020, Molinaro *et al.* 2021, Browning *et al.* 2023). Of these methods, RFMix v.1.5.4 is considered to be state-of-the-art in estimating ancestry in complex admixture scenarios (Daya *et al.* 2014, Uren *et al.* 2020, Hilmarsson *et al.* 2021). Furthermore, an updated version of this software—called RFMIX version 2 (RFMIX2)

(<https://github.com/slowkoni/rfmix>)—was recently developed and is argued to be superior to the previous version (Carrot-Zhang *et al.* 2021). However, there is currently no straightforward way to visualize the global and local ancestry results from RFMIX2, which is important for the exploration, interpretation, and reporting of ancestry patterns in human populations. To address this problem, we developed computational tools to depict both global and local ancestry, as inferred by RFMIX2, in individuals from admixed groups.

2 Implementation

Inferences of genetic ancestry are informative for determining the population origins of alleles associated with traits, including disease susceptibility (Freedman *et al.* 2006, Cheng *et al.* 2009, Daya *et al.* 2014) and for understanding the genetic history of admixed populations (most notably, past migration events) (Daya *et al.* 2014, Uren *et al.* 2020; Gopalan *et al.* 2022; Browning *et al.* 2023).

The AncestryGrapher toolkit is a very powerful package for displaying the broad ancestry proportions originating from source populations that individuals possess (i.e. global ancestry), which is helpful when selecting cohorts of individuals with either a high or low level of a particular ancestry for admixture mapping analyses (Caliebe *et al.* 2022; Shriner 2023). Furthermore, the AncestryGrapher toolkit can visualize the genetic ancestry within genomic coordinates along chromosomes, allowing for the efficient identification of the population origins of alleles (i.e. local ancestry) in anthropological studies (Gravel 2012, Secolin *et al.* 2019). The AncestryGrapher toolkit enables users to visualize these levels of ancestry with two distinct pipelines, Global Ancestry Painting (GAP) and Local Ancestry Painting (LAP; Fig. 1).

2.1 Global Ancestry Painting

The Global Ancestry Painting (GAP) pipeline consists of three separate Python scripts: (i) RFMIX2ToBed4GAP.py; (ii) BedToGAP.py; and (iii) GAP.py.

2.1.1 Step 1: Combine the RFMIX2 output files for all chromosomes and individuals into a single file

The RFMIX2 software generates a *.rfmix.Q (global ancestry information) output file for each chromosome per individual in a given dataset. The RFMIX2ToBed4GAP.py script will combine these separate files into a single file and then calculate a mean value for each inferred ancestry component across chromosomes per individual. If users wish to sort individuals by a particular mean ancestry component (from largest to smallest), they can accomplish this task by implementing the “--sort-ancestry” flag.

2.1.2 Step 2: Create the input file for GAP to visualize the global ancestry components

The output file generated in Step 1 will serve as the input file for BedToGAP.py. Here, users can select up to 10 distinct colors to represent ancestry (one color for each mean ancestry component).

2.1.3 Step 3: Generate the plot for global ancestry components

In this step, users will run the GAP.py script to generate an output file with mean ancestry components in a bar plot for all individuals, together with a legend of ancestry origin and

the name of each individual in the dataset, in either “pdf” or “svg” format (Fig. 1A).

2.2 Local Ancestry Painting

The Local Ancestry Painting (LAP) pipeline consists of three separate Python scripts: (i) RFMIX2ToBed.py; (ii) BedToLAP.py; and (iii) LAP.py.

2.2.1 Step 1: Combine the output files from RFMIX2 into a single file and generate *.bed files

RFMIX2 generates a *.msp.tsv (local ancestry information) output file for each chromosome (e.g. 1–22) for a given individual. RFMIX2ToBed.py will combine these chromosomes into a single file that contains header and ancestry information for all chromosomes for each individual in the dataset. RFMIX2ToBed.py will also generate two *.bed files (*_hap1.bed and *_hap2.bed) per individual, corresponding to maternal and paternal chromosomes in diploid organisms (with one set of chromosomes in each *.bed file).

2.2.2 Step 2: Create the color scheme for ancestry painting along chromosomes

Here, the output files from Step 1 will serve as the input files in this step. BedToLAP.py will assign a default color to each ancestry component inferred by RFMIX2 (a maximum of 10 colors will be generated, one for each ancestry component). Alternatively, users can select up to 10 distinct colors with the “--ancestry” flag. BedToLAP.py can also highlight a specific gene or genomic region of interest on a chromosome with additional parameters.

2.2.3 Step 3: Generate the ancestry plot for each chromosome for a given individual

The resulting output files, generated using the LAP.py script, will contain ancestry-informative karyograms along with a legend of ancestry origin and other identifiers (Fig. 1B). The images within the output files will have 4k resolution (4210 × 1663) and can be edited in Adobe Illustrator or Inkscape.

3 Results and discussion

We have developed the AncestryGrapher toolkit to visually summarize the output of the RFMIX2 software. To demonstrate the utility of our computational pipelines, we applied the GAP and LAP algorithms to plot the inferred ancestry of individuals in the North African Mozabite Berber population from the Human Genome Diversity Panel (Rosenberg *et al.* 2002, Li *et al.* 2008, Bergström *et al.* 2020). Our results showed that individuals in this population have distinct levels of Middle Eastern, European, and sub-Saharan African ancestry (Fig. 1), consistent with prior studies (Henn *et al.* 2012, Bekada *et al.* 2015, Arauna *et al.* 2017, Silva *et al.* 2021).

In addition, we detected striking patterns of local ancestry along individual chromosomes. For this analysis, we focused our attention on the well-studied regulatory region in intron 13 of the MCM6 gene on Chromosome 2, which contains single nucleotide polymorphisms associated with lactase persistence—namely, C/T₋₁₃₉₁₀ (rs4988235), C/G₋₁₃₉₀₇ (rs41525747), T/G₋₁₃₉₁₅ (rs41380347), T/G₋₁₄₀₀₉ (rs869051967), and G/C₋₁₄₀₁₀ (rs145946881) (Ennaffah *et al.* 2002, Ingram *et al.* 2007, 2009, Tishkoff *et al.* 2007, Hassan *et al.* 2016, Anguita-Ruiz *et al.* 2020, Campbell and Ranciaro 2021). Prior studies also have shown that these functional sites have been the targets of Darwinian selection in Middle Eastern, European, and sub-

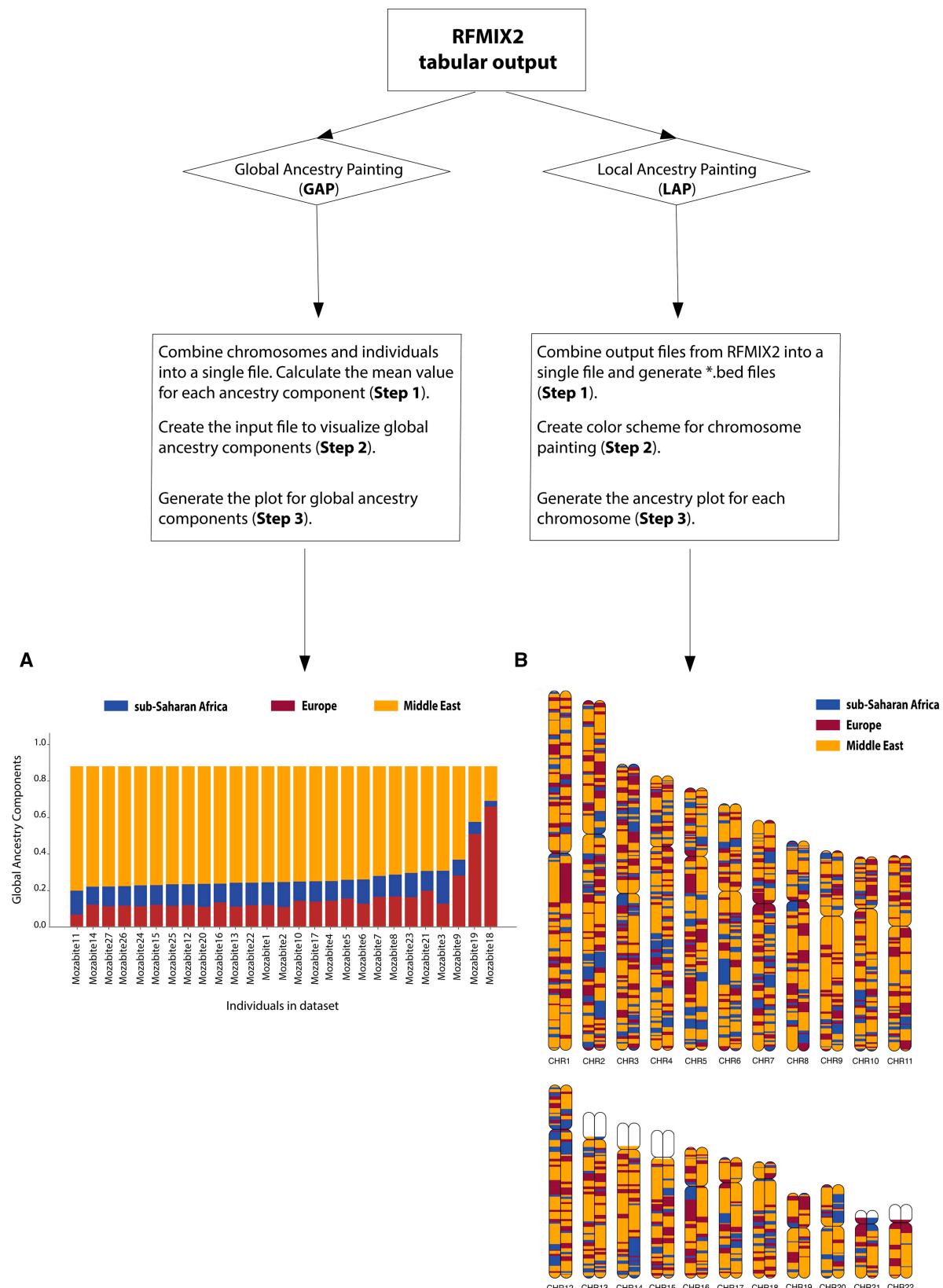


Figure 1. Workflow of the AncestryGrapher toolkit. AncestryGrapher consists of two pipelines: (i) GAP and (ii) LAP that run in a command-line Terminal. GAP combines individuals and chromosomes into a single file, calculates mean ancestry components across chromosomes per individual, and visualizes their ancestry in a bar plot (i.e., global ancestry). Panel A shows the plotted mean ancestry components per individual in the Mozabite population from the Human Genome Diversity Panel (Rosenberg *et al.* 2002; Li *et al.* 2008; Bergström *et al.* 2020). Individuals have been arranged by the Middle Eastern ancestry component from largest to smallest. LAP combines the output from RFMIX2 into a single file, creates a color scheme for ancestry painting, and plots the ancestry along chromosomes (i.e., local ancestry). Panel B shows the local ancestry components along diploid Chromosomes 1 through 22 for a Mozabite individual. In Panels A and B, the orange color indicates Middle Eastern ancestry; purple represents European ancestry; and blue signifies sub-Saharan African ancestry. The white color in Panel B represents regions of the genome with unknown ancestry (i.e., the reference ancestry is not present) and/or these regions contain missing SNP data.

Saharan African populations with a history of dairying (Bersaglieri *et al.* 2004, Tishkoff *et al.* 2007, Ennaffah *et al.* 2008, Gerbault *et al.* 2009, Itan *et al.* 2009, Jones *et al.* 2013, Macholdt *et al.* 2014, Ranciaro *et al.* 2014, Scheinfeldt *et al.* 2019, Vicente *et al.* 2019, Hollfelder *et al.* 2020). Based on our LAP plots, we found that alleles in the *MCM6* regulatory sequence in the Mozabite Berber population sit on a chromosomal segment on Chromosome 2 originating from the Middle East, Europe, and/or sub-Saharan Africa. More specifically, alleles in the *MCM6* regulatory region were found to sit in a chromosomal segment originating from the Middle East on 53.7% of chromosomes, while alleles in the same region had ancestry backgrounds originating from Europe and sub-Saharan Africa on 31.5% and 14.8% of chromosomes, respectively, in the Mozabite Berbers (Supplementary Fig. S1). Thus, genetic variation in the *MCM6* regulatory region was likely introduced into this population through admixture, which we visually captured using the LAP pipeline. For comparison, we also show GAP and LAP results for different populations with a history of geographic and cultural isolation, namely the Finnish (Europe) and the Bedouin (Middle East) in Supplementary Figs S2 and S3, respectively.

In summary, the AncestryGrapher toolkit has the capability to calculate the mean ancestry components based on RFMIX2 output files and provide a graphical display of this global ancestry for each individual in a given dataset. Furthermore, this package can convert the ancestry tracts from RFMIX2 into colors representing ancestry along chromosomes. While these strengths make AncestryGrapher a powerful tool for exploring, interpreting, and reporting results, it is also important to note the limitations of our approaches and the RFMIX2 software. Although there is no restriction on the number of target individuals that can be visualized with the AncestryGrapher toolkit, the GAP and LAP pipelines can at most assign colors to a maximum of 10 ancestry components (or source populations). Consequently, ancestry components above the 10th one will not be displayed. Regarding RFMIX2, this software is especially effective when users have included populations that have contributed genetic ancestry to the admixed population of interest. In addition, these RFMIX2 analyses are powered by well-characterized genetic maps for SNP loci along chromosomes. Thus, RFMIX2 should only be applied to datasets with valid reference populations based on *a priori* information and reliable genetic maps. Assuming RFMIX2 is used correctly, the AncestryGrapher toolkit is highly beneficial for graphically summarizing genetic patterns in complex populations supporting anthropological and biomedical research (for more detailed information about RFMIX2 and other ancestry methods, please see Supplementary Table S1).

Acknowledgements

We thank the Center for Advanced Research Computing (CARC) at the University of Southern California for providing the computational resources for this project. We also thank our novice and experienced users for testing the reproducibility of our computational pipelines.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported in part by a National Science Foundation grant SBE-BCS 2221920 to M.C.C.

Data availability

The data underlying this article are available in the Human Genome Diversity Panel at https://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/statphase/ and in the 1000 Genomes Consortium Project at <https://www.internationalgenome.org/data-portal/population/FIN>.

References

1000 Genomes Project Consortium; Abecasis GR, Auton A, Brooks LD *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56–65.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655–64.

Anguita-Ruiz A, Aguilera CM, Gil Á. Genetics of lactose intolerance: an updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients* 2020;12:2689.

Arauna LR, Mendoza-Revilla J, Mas-Sandoval A *et al.* Recent historical migrations have shaped the gene pool of Arabs and Berbers in North Africa. *Mol Biol Evol* 2017;34:318–29.

Baran Y, Pasaniuc B, Sankararaman S *et al.* Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* 2012;28:1359–67.

Bekada A, Arauna LR, Deba T *et al.* Genetic heterogeneity in Algerian human populations. *PLoS One* 2015;10:e0138453.

Bergström A, McCarthy SA, Hui R *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* 2020;367:eaay5012.

Bersaglieri T, Sabeti PC, Patterson N *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004;74:1111–20.

Brisbin A, Bryc K, Byrnes J *et al.* PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 2012;84:343–64.

Browning SR, Browning BL, Zhou Y *et al.* Analysis of human sequence data reveals two pulses of archaic Denisovan admixture. *Cell* 2018;173:53–61.e9.

Browning SR, Waples RK, Browning BL. Fast, accurate local ancestry inference with FLARE. *Am J Hum Genet* 2023;110:326–35.

Bryc K, Auton A, Nelson MR *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci U S A* 2010;107:786–91.

Caliebe A, Tekola-Ayele F, Darst BF *et al.*; IGES ELSI Committee. Including diverse and admixed populations in genetic epidemiology research. *Genet Epidemiol* 2022;46:347–71.

Campbell MC, Hirbo JB, Townsend JP *et al.* The peopling of the African continent and the diaspora into the new world. *Curr Opin Genet Dev* 2014;29:120–32.

Campbell MC, Ranciaro A. Human adaptation, demography and cattle domestication: an overview of the complexity of lactase persistence in Africa. *Hum Mol Genet* 2021;30:R98–R109.

Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 2008;9:403–33.

Carrot-Zhang J, Han S, Zhou W *et al.* Analytical protocol to identify local ancestry-associated molecular features in cancer. *STAR Protoc* 2021;2:100766.

Cheng C-Y, Kao WHL, Patterson N *et al.* Admixture mapping of 15,280 african americans identifies obesity susceptibility loci on chromosomes 5 and X. *PLoS Genet* 2009;5:e1000490.

Churchhouse C, Marchini J. Multiway admixture deconvolution using phased or unphased ancestral panels. *Genet Epidemiol* 2013; 37:1–12.

Daya M, van der Merwe L, Gignoux CR *et al.* Using multi-way admixture mapping to elucidate TB susceptibility in the South African Coloured population. *BMC Genomics* 2014;15:1021.

Dias-Alves T, Mairal J, Blum MGB. Loter: a software package to infer local ancestry for a wide range of species. *Mol Biol Evol* 2018; 35:2318–26.

Durand EY, Do CB, Mountain JL *et al.* Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. bioRxiv, <https://doi.org/10.1101/010512>, 2014, preprint: not peer reviewed.

Enattah NS, Jensen TGK, Nielsen M *et al.* Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 2008; 82:57–72.

Enattah NS, Sahl T, Savilahti E *et al.* Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002;30:233–7.

Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 2003;164:1567–87.

Freedman ML, Haiman CA, Patterson N *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 2006;103:14068–73.

Gerbault P, Moret C, Currat M *et al.* Impact of selection and demography on the diffusion of lactase persistence. *PLoS One* 2009; 4:e6369.

Goli RC, Chishi KG, Ganguly I *et al.* Global and local ancestry and its importance: a review. *Curr Genomics* 2024;25:237–60.

Gopalan S, Smith SP, Korunes K *et al.* Human genetic admixture through the lens of population genomics. *Philos Trans R Soc Lond B Biol Sci* 2022;377:20200410.

Gravel S. Population genetics models of local ancestry. *Genetics* 2012; 191:607–19.

Guan Y. Detecting structure of haplotypes and local ancestry. *Genetics* 2014;196:625–42.

Gurdasani D, Carstensen T, Tekola-Ayele F *et al.* The African Genome Variation Project shapes medical genetics in Africa. *Nature* 2015; 517:327–32.

Hassan HY, van Erp A, Jaeger M *et al.* Genetic diversity of lactase persistence in East African populations. *BMC Res Notes* 2016;9:8.

Hellenthal G, Busby GB, Band G *et al.* A genetic atlas of human admixture history. *Science* 2014;343:747–51.

Henn BM, Botigué LR, Gravel S *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet* 2012; 8:e1002397.

Hilmarsdóttir H, Kumar AS, Rastogi R *et al.* High resolution ancestry deconvolution for next generation genomic data. bioRxiv, <https://doi.org/10.1101/2021.09.19.460980>, 2021, preprint: not peer reviewed.

Hoggart CJ, Shriver MD, Kittles RA *et al.* Design and analysis of admixture mapping studies. *Am J Hum Genet* 2004;74:965–78.

Hollfelder N, Babiker H, Granehäll L *et al.* The genetic variation of lactase persistence alleles in northeast Africa. bioRxiv, <https://doi.org/10.1101/2020.04.23.205735>, 2020, preprint: not peer reviewed.

Ingram CJE, Elamin MF, Mulcare CA *et al.* A novel polymorphism associated with lactose tolerance in Africa: multiple causes for lactase persistence? *Hum Genet* 2007;120:779–88.

Ingram CJE, Raga TO, Tarekegn A *et al.* Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. *J Mol Evol* 2009; 69:579–88.

Itan Y, Powell A, Beaumont MA *et al.* The origins of lactase persistence in Europe. *PLoS Comput Biol* 2009;5:e1000491.

Jones BL, Raga TO, Liebert A *et al.* Diversity of lactase persistence alleles in Ethiopia: signature of a soft selective sweep. *Am J Hum Genet* 2013;93:538–44.

Koller D, Wendt FR, Pathak GA *et al.* Denisovan and Neanderthal archaic introgression differentially impacted the genetics of complex traits in modern populations. *BMC Biol* 2022;20:249.

Korunes KL, Goldberg A. Human genetic admixture. *PLoS Genet* 2021;17:e1009374.

Lawson DJ, Hellenthal G, Myers S *et al.* Inference of population structure using dense haplotype data. *PLoS Genet* 2012;8:e1002453.

Li JZ, Absher DM, Tang H *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; 319:1100–4.

Liu Y, Nyunoya T, Leng S *et al.* Softwares and methods for estimating genetic ancestry in human populations. *Hum Genomics* 2013;7:1.

Macholdt E, Lede V, Barbieri C *et al.* Tracing pastoralist migrations to Southern Africa with lactase persistence alleles. *Curr Biol* 2014; 24:875–9.

Maples BK, Gravel S, Kenny EE *et al.* RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 2013;93:278–88.

Molinaro L, Marnetto D, Mondal M *et al.* A chromosome-painting-based pipeline to infer local ancestry under limited source availability. *Genome Biol Evol* 2021;13:evab025.

Montserrat DM, Bustamante C, Ioannidis A. Lai-net: local-ancestry inference with neural networks. In: *ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 2020, 1314–18.

Pasaniuc B, Sankararaman S, Kimmel G *et al.* Inference of locus-specific ancestry in closely related populations. *Bioinformatics* 2009; 25:i213–21.

Patterson N, Hattangadi N, Lane B *et al.* Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; 74:979–1000.

Price AL, Tandon A, Patterson N *et al.* Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 2009;5:e1000519.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.

Qin P, Stoneking M. Denisovan ancestry in east Eurasian and Native American populations. *Mol Biol Evol* 2015;32:2665–74.

Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 2014; 197:573–89.

Ranciaro A, Campbell MC, Hirbo JB *et al.* Genetic origins of lactase persistence and the spread of pastoralism in Africa. *Am J Hum Genet* 2014;94:496–510.

Rosenberg NA, Pritchard JK, Weber JL *et al.* Genetic structure of human populations. *Science* 2002;298:2381–5.

Rudan I. Health effects of human population isolation and admixture. *Croat Med J* 2006;47:526–31.

Salter-Townshend M, Myers S. Fine-scale inference of ancestry segments without prior knowledge of admixing groups. *Genetics* 2019; 212:869–89.

Sánchez-Quinto F, Lalueza-Fox C. Almost 20 years of Neanderthal palaeogenetics: adaptation, admixture, diversity, demography and extinction. *Philos Trans R Soc Lond B Biol Sci* 2015;370: 20130374.

Sankararaman S, Kimmel G, Halperin E *et al.* On the inference of ancestries in admixed populations. *Genome Res* 2008a;18:668–75.

Sankararaman S, Mallick S, Patterson N *et al.* The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Curr Biol* 2016;26:1241–7.

Sankararaman S, Sridhar S, Kimmel G *et al.* Estimating local ancestry in admixed populations. *Am J Hum Genet* 2008b;82:290–303.

Scheinfeldt LB, Soi S, Lambert C *et al.* Genomic evidence for shared common ancestry of East African hunting-gathering populations and insights into local adaptation. *Proc Natl Acad Sci U S A* 2019; 116:4166–75.

Secolini R, Mas-Sandoval A, Arauna LR *et al.* Distribution of local ancestry and evidence of adaptation in admixed populations. *Sci Rep* 2019;9:13900.

Shriner D. Overview of admixture mapping. *Curr Protoc* 2023;3:e677.

Silva M, Oteo-García G, Martiniano R *et al.* Biomolecular insights into North African-related ancestry, mobility and diet in eleventh-century Al-Andalus. *Sci Rep* 2021;11:18121.

Sundquist A, Fratkin E, Do CB *et al.* Effect of genetic divergence in identifying ancestral origin using HAPAA. *Genome Res* 2008a; 18:676–82.

Sundquist A, Fratkin E, Do CB *et al.* Effects of genetic divergence in identifying ancestral origin using HAPAA. In: Vingron M, Wong, L (eds), *Research in Computational Molecular Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008b, 423.

Tan T, Atkinson EG. Strategies for the genomic analysis of admixed populations. *Annu Rev Biomed Data Sci* 2023;6:105–27.

Tang H, Coram M, Wang P *et al.* Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet* 2006;79:1–12.

Tang H, Peng J, Wang P *et al.* Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol* 2005; 28:289–301.

Tishkoff SA, Reed FA, Ranciaro A *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 2007; 39:31–40.

Uren C, Hoal EG, Möller M. Putting RFMix and ADMIXTURE to the test in a complex admixed population. *BMC Genet* 2020;21:40.

Uren C, Kim M, Martin AR *et al.* Fine-scale human population structure in Southern Africa reflects ecogeographic boundaries. *Genetics* 2016;204:303–14.

Vespasiani DM, Jacobs GS, Cook LE *et al.* Denisovan introgression has shaped the immune system of present-day Papuans. *PLoS Genet* 2022;18:e1010470.

Vicente M, Priezodová E, Diallo I *et al.* Population history and genetic adaptation of the Fulani nomads: inferences from genome-wide data and the lactase persistence trait. *BMC Genomics* 2019; 20:915.

Vyas DN, Mulligan CJ. Analyses of Neanderthal introgression suggest that Levantine and Southern Arabian populations have a shared population history. *Am J Phys Anthropol* 2019;169: 227–39.

Wall JD, Yoshihara Caldeira Brandt D. Archaic admixture in human history. *Curr Opin Genet Dev* 2016;41:93–7.

Wangkumhang P, Greenfield M, Hellenthal G. An efficient method to identify, date, and describe admixture events using haplotype information. *Genome Res* 2022;32:1553–64.

Witt KE, Funk A, Anorve-Garibay V *et al.* The impact of modern admixture on archaic human ancestry in human populations. *Genome Biol Evol* 2023;15:evad066.